

# COVID19-Unemployment Documentation

Ji-Eun Lee (jvlee@seas.upenn.edu)

# Overview

Time series data from [U.S. Bureau of Labor Statistics](#)

- [Seasonally adjusted](#) local area unemployment rate
- Jan 1990 to March 2020

Forecast unemployment rates for US local areas

ARIMA(p,d,q) model (AutoRegressive Integrated Moving Average)

$$\hat{y}_t = \mu + \phi_1 y_{t-1} + \dots + \phi_p y_{t-p} - \theta_1 e_{t-1} - \dots - \theta_q e_{t-q} \quad \text{If } d=0: y_t = Y_t$$

$$\text{If } d=1: y_t = Y_t - Y_{t-1}$$

$$\text{If } d=2: y_t = (Y_t - Y_{t-1}) - (Y_{t-1} - Y_{t-2})$$

# Parameter search

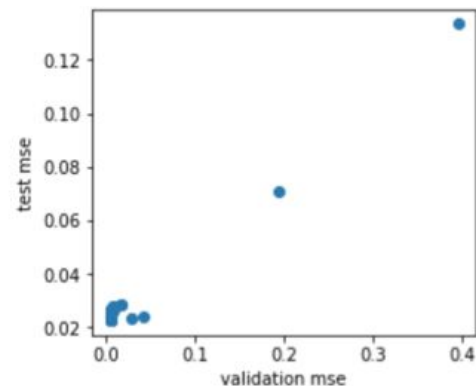
Grid search for order=(p, d, q) of ARIMA model

11x3x4 (132) orders searched

Train set [1990:2008] | Validation set [2009:2013] | Test set [2014:2020]

The validation mse and test mse shows proportional relationship

- Safe to use the parameter



# Directory structure

Hershey DILab Google Drive > Pennovation > COVID19-Unemployment

- dataset/
  - area.txt
  - states/
- results/
- unemployment.ipynb

If directory location is changed, update the 3rd cell in Part 1 of unemployment.ipynb

```
[3] 1 cd drive/My\ Drive/Pennovation/COVID19-Unemployment
```

```
↳ /content/drive/My Drive/Pennovation/COVID19-Unemployment
```

# Code






Python, Jupyter

- Run using Google Colab
- Jupyter to run locally

Make sure to run all cells sequentially to have all the necessary functions and variables defined

Daily threshold of API calls per user

- Approximately can fetch data for six states per day
- dataset/states: state level data for Jan 1990 - Mar 2020
  - Needs to be updated manually (using ``data_update`` function)
    - Run ``data_update`` before using the retrieved time series
    - This will update the time series and the saved 'STATE.txt'
    - Refer to Line 4 of Part 4

My Drive > ... > dataset > state ▾		
Name ↑		Owner
 Alabama.txt		me
 Alaska.txt		me
 Arizona.txt		me
 Arkansas.txt		me
 California.txt		me

# How to run

## [Part 4] Run

```
1 area_text = "" # STATE NAME
2 curr_year = 2020 # CURRENT YEAR
3
4 ts = data_update(state_data[area_text], curr_year, area_text)
5 results = grid_search(ts)
6
7 for i in range(3):
8     print('ARIMA%s \t Validation MSE=%.3f \t Test MSE=%.3f' % (results[i][0], results[i][1][0], results[i][1][1]))
9     forecast(ts, area_text, order=results[i][0], future_steps=12*5)
10    check_forecast(ts, area_text, list(range(1990,2009)), order=results[i][0], future_steps=12*5)
```

Line 5: The output of grid\_search is a list of tuples

- Each tuple is (order, (validation mse, test mse))
- List is sorted by lowest validation mse
- Plots the validation mse - test mse relation

Line 7-10: Shows the results using the orders with the first, second, and third lowest validation mse

**Line 9:** Shows the forecast of the next 5 years when the model uses [1990, Mar 2020] to train

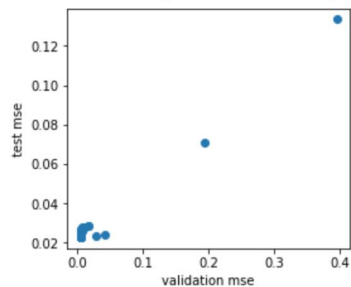
Line 10: Shows the forecast of the next 5 years when the model uses [1990, 2008] to train

- Can observe that the forecast diverges after the second year
- "Little less than a year is 'reliable', year 2 is 'directionally correct' and after that is a guess"

*Refer to Part 5 to see how the execution of this cell should look like*

# Sample execution: Pennsylvania

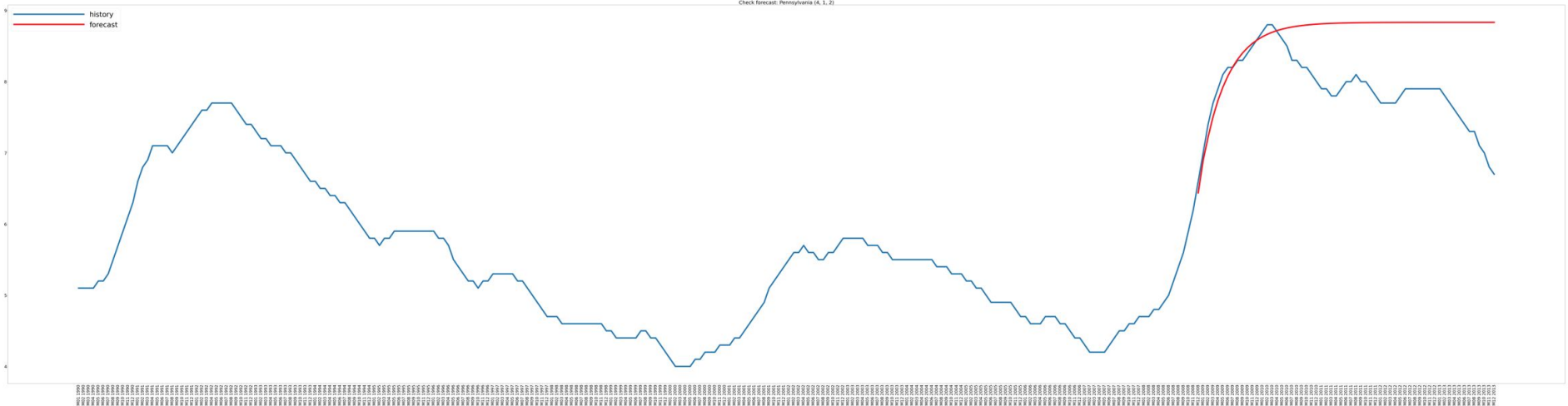
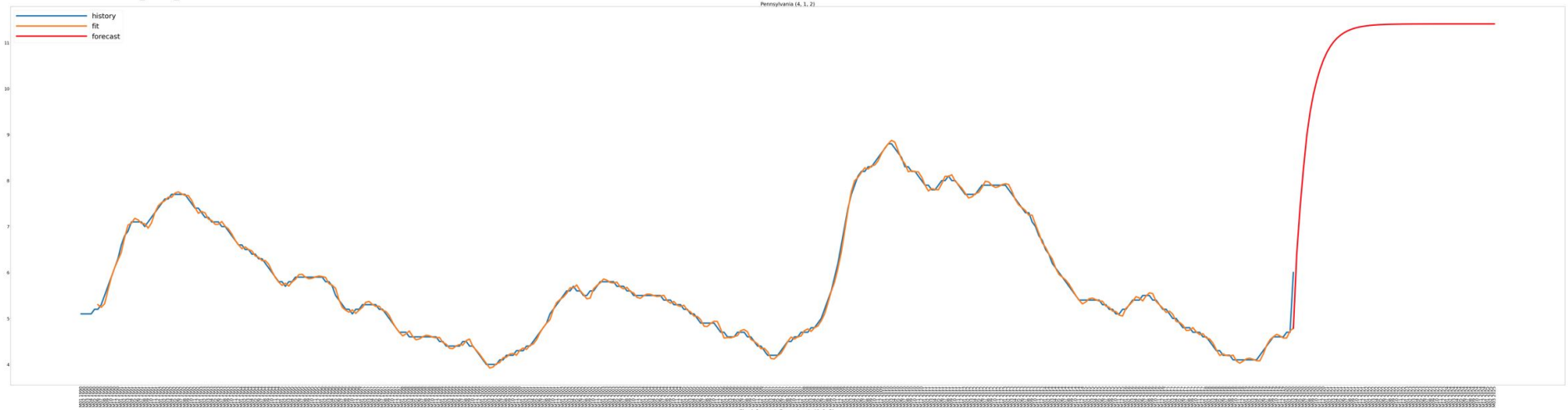
```
ARIMA(9, 1, 0) Validation MSE=0.005 Test MSE=0.023
ARIMA(9, 1, 1) Validation MSE=0.005 Test MSE=0.023
ARIMA(9, 1, 2) Validation MSE=0.005 Test MSE=0.023
ARIMA(9, 1, 3) Validation MSE=0.005 Test MSE=0.023
ARIMA(9, 2, 0) Validation MSE=0.005 Test MSE=0.023
ARIMA(9, 2, 1) Validation MSE=0.005 Test MSE=0.023
ARIMA(9, 2, 2) Validation MSE=0.005 Test MSE=0.023
ARIMA(9, 2, 3) Validation MSE=0.005 Test MSE=0.023
ARIMA(10, 0, 0) Validation MSE=0.005 Test MSE=0.023
ARIMA(10, 0, 1) Validation MSE=0.005 Test MSE=0.023
ARIMA(10, 0, 2) Validation MSE=0.005 Test MSE=0.023
ARIMA(10, 0, 3) Validation MSE=0.004 Test MSE=0.023
ARIMA(10, 1, 0) Validation MSE=0.005 Test MSE=0.023
ARIMA(10, 1, 1) Validation MSE=0.005 Test MSE=0.023
ARIMA(10, 1, 2) Validation MSE=0.005 Test MSE=0.023
ARIMA(10, 1, 3) Validation MSE=0.004 Test MSE=0.023
ARIMA(10, 2, 0) Validation MSE=0.005 Test MSE=0.023
ARIMA(10, 2, 1) Validation MSE=0.005 Test MSE=0.023
ARIMA(10, 2, 2) Validation MSE=0.005 Test MSE=0.023
ARIMA(10, 2, 3) Validation MSE=0.005 Test MSE=0.023
Grid search completed. Time elapsed 7001.03 seconds
```



validation mse - test mse: proportional relationship

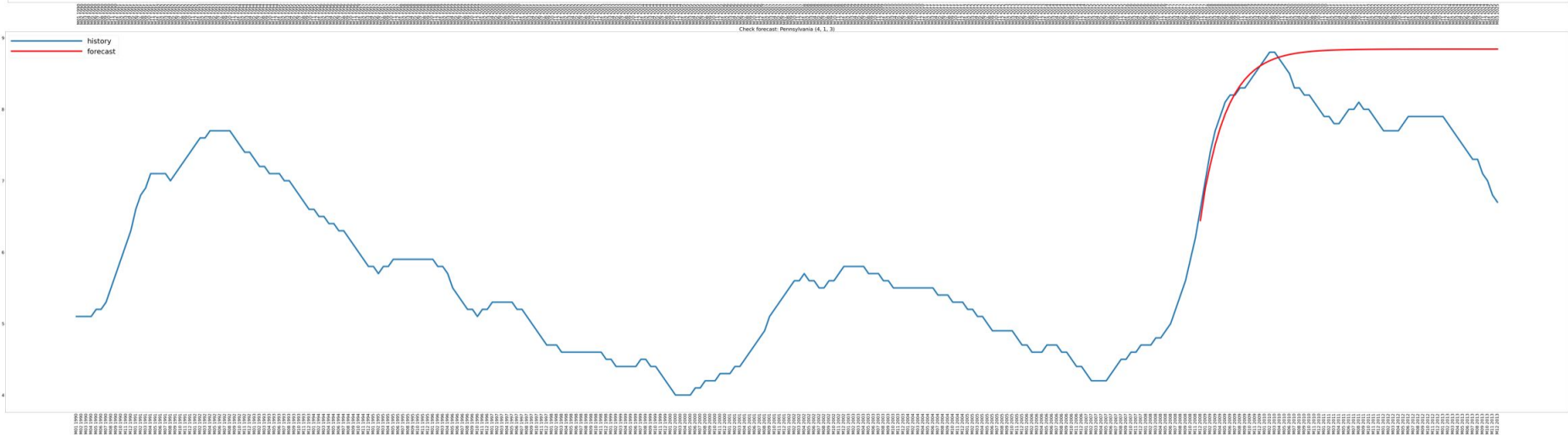
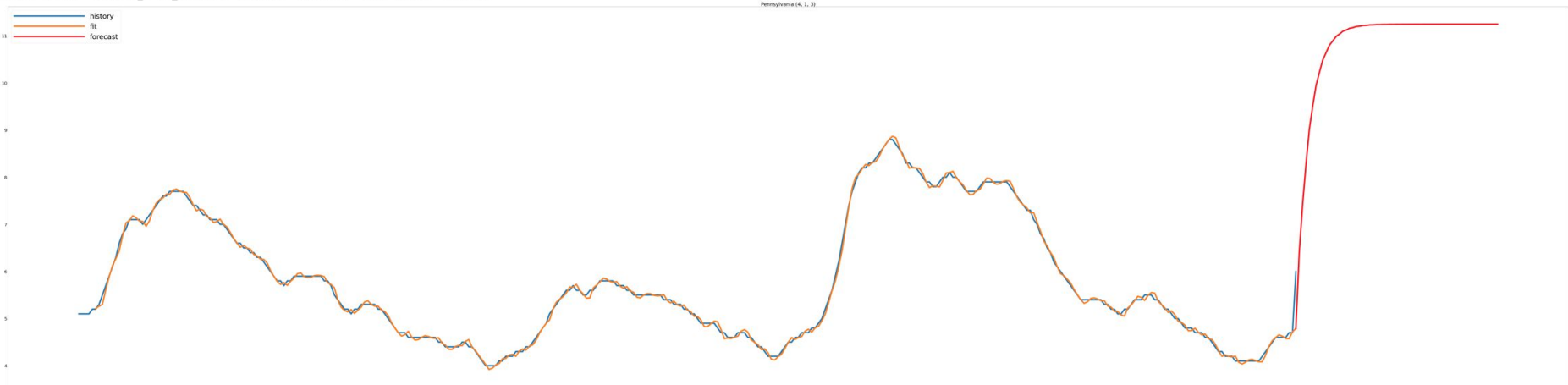
Best model: ARIMA(4, 1, 2) Validation\_MSE=0.004 Test MSE=0.024

[1] ARIMA(4, 1, 2)      Validation MSE=0.004      Test MSE=0.024  
Maximum unemployment rate 11.41 at M03 2025





[2] ARIMA(4, 1, 3)      Validation MSE=0.004      Test MSE=0.024  
Maximum unemployment rate 11.25 at M11 2023



[3] ARIMA(2, 0, 3)      Validation MSE=0.004      Test MSE=0.023  
Maximum unemployment rate 10.51 at M02 2021

